

Beyond the boundaries of content: transparency on training data and the licensing market in generative AI in the European Union: between innovation and holders protection

Federico Domenico Enrico De Silvo

"Parthenope" University of Naples, Italy
federico.desilvo@uniparthenope.it-

Abstract

The paper argues that transparency into the content used to train general-purpose artificial intelligence models operates, under European Union law, as a market infrastructure capable of reducing information asymmetries and transaction costs, making the transition from contention to licensing more viable. Moving from the “conditional” regime of text-and-data mining in Directive (EU) 2019/790 – based on lawful access and, according to Article 4, on the reservation of holders – the analysis aims to show how standardized disclosure (Commission 2025 template) strengthens the economic/legal verifiability of claims and incentives for licensing agreements, especially for high-quality content. Based, therefore, on institutional and doctrinal evidence, the conditions and limits of such “market-making transparency” (opt-out effectiveness, information granularity, compensation sustainability) are discussed, thus outlining implications for rights governance and contractual strategies along the content chain.

Keywords: Training-data transparency, Generative AI (GPAI) governance, EU AI Act compliance, Text and data mining (DSM Directive), Rights reservation (opt-out) and enforcement, Licensing markets and transaction costs, Knowledge governance and organisational change